IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s): Y-C. Chang et al.
Docket No.:   YOR920030366US1
Serial No.:   10/668,549
Filing Date: September 23, 2003
Group: 2161
Examiner: Monica M. Pyo

Title: Methods and Apparatus for Query Rewrite with
           Auxiliary Attributes in Query Processing Operations

---

APPEAL BRIEF

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

     Applicants (hereinafter referred to as "Appellants") hereby appeal the final rejection, dated August 1, 2008, of claims 1-4, 7-10 and 21 of the above-identified application.

REAL PARTY IN INTEREST

     The present application is assigned to International Business Machines Corporation. The assignee, International Business Machines Corporation, is the real party in interest.

RELATED APPEALS AND INTERFERENCES

     There are no known related appeals or interferences.

STATUS OF CLAIMS

The present application was filed on September 23, 2003 with claims 1-21. Claims 5, 6 and 11-20 have been previously canceled. Claims 1-4, 7-10 and 21 remain pending and claim 1 is the only pending independent claim.

Claims 1-3, 7, 9, 10, and 21 are finally rejected under 35 U.S.C. §102(e). Claims 4 and 8 are finally rejected under 35 U.S.C. §103(a). Claims 1-4, 7-10 and 21 are appealed.

STATUS OF AMENDMENTS

There have been no amendments filed subsequent to the final rejection.

SUMMARY OF CLAIMED SUBJECT MATTER

Independent claim 1 is directed to a method of rewriting a query during a database query processing operation. The method includes a step of processing the query having one or more target attributes in accordance with at least a portion of a data set producing query results comprising the one or more target attributes and one or more auxiliary attributes. The one or more auxiliary attributes are not included in the query. The method also includes a step of analyzing the query results with respect to the one or more target attributes and the one or more auxiliary attributes to determine a relative selectivity for the one or more auxiliary attributes. The method includes a further step of appending the query with at least one new predicate corresponding to at least one of the one or more auxiliary attributes having a high relative selectivity to form a rewritten query.

As described in the specification at, for example, page 4, lines 9-15, page 5, lines 15-25, and page 6, lines 23-25, an illustrative method (e.g., FIGS. 2 and 5) of rewriting a query (e.g., 200 in FIG. 2 and 508 in FIG. 5) during a database query processing operation includes a step (e.g., 202 in FIG. 2) of processing the query having one or more target attributes in accordance with at least a portion of a data set producing query results comprising the one or more target attributes and one or more auxiliary attributes which are not included in the query. As described in the specification at, for example, page 6, line 18, to page 7, line 3, with reference to FIG. 4, the method also includes a step (e.g., 212 in FIG. 2) analyzing the query results with respect to the one or more target attributes and

the one or more auxiliary attributes to determine a relative selectivity for the one or more auxiliary attributes. As described in the specification at, for example, page 7, lines 4-22, with reference to FIG. 5, the method includes a further step (e.g., 506 in FIG. 5) of appending the query (e.g., 508 in FIG. 5) with at least one new predicate (e.g., 504 in FIG. 5) corresponding to at least one of the one or more auxiliary attributes having a high relative selectivity (e.g., 502 in FIG. 5) to form a rewritten query (e.g., 510 in FIG. 5).

Illustrative embodiments of the present invention provide a number of significant advantages over conventional techniques. For example, as discussed in the specification at page 2, line 16, to page 3, line 7, illustrative embodiments may: (i) may advocate a query rewrite strategy to append new predicates based on attributes absent from the original query; (ii) may employ a two-pass query processing operation approach as opposed to the prevalent single-pass processing; (iii) may employ statistical analysis and other data mining techniques to help identify distinguishing attributes; and (iv) may have false dismissals due to query rewrite, while traditional methodologies emphasize exact answers. Thus, illustrative embodiments leverage the knowledge about auxiliary attributes in data records to append additional query predicates for more efficient processing. Advantageously, such illustrative embodiments may be useful for providing efficient query processing operations in large databases, while maintaining the accuracy of results. This efficiency translates a normally high processing cost query into a low cost query.

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

1. Claims 1-3, 7, 9, 10, and 21 are rejected under 35 U.S.C. §102(e) as being anticipated by U.S. Patent No. 6,581,056 (hereinafter "Rao").

2. Claim 4 is rejected under 35 U.S.C. §103(a) as being unpatentable over Rao in view of U.S. Patent Application Publication No. 2003/0167259 (hereinafter "Casson").

3. Claim 8 is rejected under 35 U.S.C. §103(a) as being unpatentable over Rao in view of U.S. Patent No. 5,890,150 (hereinafter "Ushijima").

3

ARGUMENT

1. Rejection of claims 1-3, 7, 9, 10, and 21 under §102(e) as being anticipated by Rao.

Claims 1, 9, 10, and 21

With regard to the §102(e) rejection of claim 1, Appellants initially note that the Examiner claims that "[i]t should be noted that a reference may be relied upon for all it would have been reasonably suggested to one having ordinary skill in the art, including nonpreferred embodiments," citing by MPEP 2123.

Appellants respectfully submit that *Merck & Co. v. Biocraft Laboratories*, 874 F.2d 804, 10 USPQ2d 1843 (Fed. Cir.), *cert. denied,* 493 U.S. 975 (1989), cited by MPEP 2123, in fact states (with citations omitted):

> Unlike a section 102 defense which requires that a single reference describe each and every element of a claimed invention, the question under 35 USC 103 is not merely what the references expressly teach but what they would have suggested to one of ordinary skill in the art at the time the invention was made.

Accordingly, Appellants respectfully assert that, as was recently reiterated by the Federal Circuit, "unless a reference discloses within the four corners of the document not only all of the limitations claimed but also all of the limitations arranged or combined in the same way as recited in the claim, it cannot be said to prove prior invention of the thing claimed and, thus, cannot anticipate under 35 U.S.C. §102." *Net MoneyIN Inc. v. VeriSign Inc.*, 545 F.3d 1359, 1369, 88 USPQ2d 1751, 1760 (Fed. Cir. 2008).

In the final Office Action, at page 7, second paragraph, the Examiner argues, with reference to Rao at column 6, line 60, to column 7, line 31, that "one or [sic] ordinary skill in the art would clearly recognize that these teachings of Rao are equivalent to the claimed feature of analyzing the query results with respect to the terms (either used in the query or not exact cluster terms) in order to refine a query."

Appellants respectfully submit that the Examiner has mischaracterized the limitation of at issue, which does not recite "analyzing the query results with respect to the terms . . . in order to

refine a query," but rather includes a limitation wherein a method of rewriting a query includes a step of analyzing the query results with respect to the one or more target attributes and the one or more auxiliary attributes to determine a relative selectivity for the one or more auxiliary attributes. Illustrative embodiments are described in the specification at, for example, page 6, line 18, to page 7, line 22, with reference to FIGS. 4 and 5.

Moreover, the relied-upon portion of Rao describes "various query refinement and browsing tools" which "allow the user to more easily utilize the information obtained by a query to formulate a more restrictive query and to evaluate higher-level strategy options," including "versions of relevance feedback, scatter/gather, snippet search, and similarity search." See Rao at column 6, lines 50-58. In other words, the relied-upon portion of Rao teaches techniques for rewriting a query based on a user's analysis of information obtained by a query.

For example, the Examiner argues that the relied-upon portion of Rao discloses "that an automatic procedure (or done by the user) is used to reweight, add or remove terms in a subsequent query." Appellants respectfully submit that the Examiner appears to be referring to Rao at column 6, lines 60-64, which states that "[i]n relevance feedback a user indicates documents in the ephemeral collection that are relevant. Typically, an automatic procedure is used to reweight, add, or remove terms in a subsequent query (although some systems permit this to be done by the user)." As such, the relevance feedback technique taught by Rao refines query terms based on the user's indication of which documents in the ephemeral collection are relevant, rather than by determining a relative selectivity for the one or more auxiliary attributes.

The Examiner also quotes from Rao at column 6, lines 65-66, which state that "[i]n scatter/gather and snippet search the focus is on 'browsing' the result set to help refine a query." However, both the scatter/gather and snippet search techniques rely on the user "browsing" the result set; see, for example, Rao at column 7, lines 13-21:

> In scatter/gather, the user can take advantage of aggregate properties and a multitude of individual properties of matching items, rather than relying on ad hoc browsing of a few items. The user can then see how the result partitions and utilize

terms from cluster term summaries or central documents to manually or automatically refine the query.

In snippet search portions of context surrounding a search term are displayable. This aids query reformulation by permitting indication of other discriminating search terms.

Again, the above-identified techniques do not involve any determination of <u>a relative selectivity</u> for the one or more auxiliary attributes.

Accordingly, Rao fails to disclose the limitation of claim 1 directed to analyzing the query results with respect to the one or more target attributes and the one or more auxiliary attributes to determine a relative selectivity for the one or more auxiliary attributes.

Dependent claims 9, 10 and 21 are patentable at least by virtue of their dependency from claim 1.

<u>Claim 2</u>

Claim 2 is patentable over the prior art of record at least by virtue of its dependency from independent claim 1, the patentability of which is discussed above. Moreover, claim 2 also defines additional, separately patentable subject matter. Specifically, claim 2 specifies that the query is processed in accordance with at least a portion of the data set, wherein the at least a portion of the data set comprises sampled records from the data set.

In illustrative embodiments discussed in the specification at, for example, page 6, lines 11-17, the use of sampled data advantageously provides a reduced number of data records for a query processing operation. The records of sampled data set may be stored in a smaller temporary database than the database which stores the full data set.

In the final Office Action at page 3, third paragraph, the Examiner argues that this limitation is met by Rao at column 4, lines 34-54, and column 4, line 65, to column 5, line 30. The relied-upon portions of Rao discusses techniques for gathering an "ephemeral collection of documents that have gathered from multiple information sources" by performing a query on these multiple data sources.

There is simply no teaching or suggestion that the query is performed on at least a portion of the data set comprising <u>sampled records</u> from the data set.

Claim 3

Claim 3 is patentable over the prior art of record at least by virtue of its dependency from claims 1 and 2, the patentability of which are discussed above. Moreover, claim 3 also defines additional, separately patentable subject matter. Specifically, claim 3 specifies that the method of claim 2 further comprises the step of sampling data records from the data set.

In the final Office Action at page 3, fourth paragraph, the Examiner argues that this limitation is met by Rao at column 6, lines 7-49. The relied-upon portion of Rao discloses that the "emphemeral SCA engine is specialized for usage on short-lived, dynamically-generated medium-sized collections that arise during interaction with information sources." As noted above with regard to claim 2, these collections arise not from a sampling operation, but rather from performing a query on these information sources. There is simply no teaching or suggestion directed to <u>sampling records</u> from a data set.

Claim 7

Claim 7 is patentable over the prior art of record at least by virtue of its dependency from independent claim 1, the patentability of which is discussed above. Moreover, claim 7 also defines additional, separately patentable subject matter. Specifically, claim 7 specifies that the step of analyzing the one or more target attributes and the one or more auxiliary attributes includes the steps of extracting statistics for the one or more auxiliary attributes from the query results, extracting statistics for the one or more auxiliary attributes from the at least a portion of the data set, and evaluating the relative selectivity for the one or more auxiliary attributes in accordance with the extracted statistics.

In the final Office Action beginning on the last paragraph of page 3, the Examiner argues that the limitation of claim 7 directed to evaluating the relative selectivity for the one or more auxiliary attributes in accordance with the extracted statistics is met by Rao at column 6, lines 18-36, which

the Examiner characterizes as teaching using "the filtering process to cause some of the tokesn [sic] to be dropped out and some tokens to be combined."

Appellants respectfully submit that the relied-upon portion of Rao states that:

> The tokenizer 502 extracts tokens by parsing the text of a document and generating a token 503 for each contiguous sequence of characters (e.g. a word). The tokens are then processed through a series of filters 504. The filters will cause some of the tokens to be dropped out and some tokens to be combined. The filters perform such functions as generating stop lists, stemming, part of speech tagging or phrase spotting. Some of the filters may also be parameterized based on the terms used in the query. In any event, after going through the series of filters 504 the remaining tokens 505 that are output are input to a statistics collector 506 for collecting the document level statistics 507 and collection level statistics 508. These statistics may include token occurrences or proximity of words or phrases.

Appellants respectfully submit that there is no teaching or suggestion directed toward evaluating the relative selectivity for the one or more auxiliary attributes, as Rao's filters are not based on the relative selectivity for one or more auxiliary attributes. Moreover, Rao specifically indicates that "after going through the series of filters 504 the remaining tokens 505 that are output are input to a statistics collector 506 for collecting the document level statistics 507 and collection level statistics 508."

In other words, the document level statistics and collection level statistics are collected based only on the tokens that remain after the filtering is performed. As such, the filtering taught by Rao clearly fails to meet the limitations of claim 1 directed to <u>evaluating</u> the relative selectivity for the one or more auxiliary attributes <u>in accordance with the extracted statistics</u>, as Rao teaches a technique the statistics are not collected until <u>after</u> the filtering is performed.

<u>2. Rejection of claim 4 under §103(a) as being unpatentable over Rao in view of Casson.</u>

Claim 4 is patentable over the prior art of record at least by virtue of its dependency from claims 1 and 3, the patentability of which are discussed above. Casson fails to remedy the above-noted deficiencies of Rao with regard to the limitations of claims 1 and 3. Accordingly, claim 4 is believed to be patentable over Rao and Casson.

3. Rejection of claim 8 under §103(a) as being unpatentable over Rao in view of Ushijima.

Claim 8 is patentable over the prior art of record at least by virtue of its dependency from claims 1 and 7, the patentability of which are discussed above. Ushijima fails to remedy the above-noted deficiencies of Rao with regard to the limitations of claims 1 and 7. Accordingly, claim 4 is believed to be patentable over Rao and Ushijima.

In view of the above, Appellants believe that claims 1-4, 7-10 and 21 are in condition for allowance, and respectfully request withdrawal of the present rejections.

Respectfully submitted,

Date: January 16, 2009

David E. Shifren
Attorney for Applicant(s)
Reg. No. 59,329
Ryan, Mason & Lewis, LLP
90 Forest Avenue
Locust Valley, NY 11560
(516) 759-2641

<u>CLAIMS APPENDIX</u>

1. A method of rewriting a query during a database query processing operation, comprising the steps of:

processing the query having one or more target attributes in accordance with at least a portion of a data set producing query results comprising the one or more target attributes and one or more auxiliary attributes, wherein the one or more auxiliary attributes are not included in the query;

analyzing the query results with respect to the one or more target attributes and the one or more auxiliary attributes to determine a relative selectivity for the one or more auxiliary attributes; and

appending the query with at least one new predicate corresponding to at least one of the one or more auxiliary attributes having a high relative selectivity to form a rewritten query.

2. The method of claim 1, wherein the at least a portion of the data set comprises sampled records from the data set.

3. The method of claim 2, further comprising the step of sampling data records from the data set.

4. The method of claim 3, wherein the step of sampling data records comprises the step of sampling every Nth record from the data set, wherein N is a positive integer.

7. The method of claim 1, wherein the step of analyzing the one or more target attributes and the one or more auxiliary attributes comprises the steps of:

extracting statistics for the one or more auxiliary attributes from the query results;

extracting statistics for the one or more auxiliary attributes from the at least a portion of the data set; and

evaluating the relative selectivity for the one or more auxiliary attributes in accordance with the extracted statistics.

8. The method of claim 7, wherein the step of evaluating the relative selectivity comprises the steps of:

comparing a range of statistics from the query results to a range of statistics from the at least a portion of the data set for the one or more auxiliary attributes; and

determining whether the one or more auxiliary attributes is a selective attribute by comparing a ratio of the ranges to a predetermined value.

9. The method of claim 1, wherein the step of appending the query with at least one new predicate comprises the steps of:

evaluating relative selectivity for the one or more auxiliary attributes;

selecting at least one auxiliary attribute with a high relative selectivity;

forming at least one new predicate; and

appending the query with the at least one new predicate.

10. The method of claim 1, further comprising the step of performing a query processing operation on a data set with a rewritten query.

21. An article of manufacture for rewriting a query during a database query processing operation; comprising a machine readable medium containing one or more programs which when executed implement the steps of claim 1.

<u>EVIDENCE APPENDIX</u>

None.

<u>RELATED PROCEEDINGS APPENDIX</u>

None.